Radiographs in the Office: Is a Second Reading Always Needed?

Paul D. Smith, MD, Jonathan Temte, MD, John W. Beasley, MD, and Marlon Mundt, MA, MS

**Background:** We evaluated the frequency, nature, and importance of the changes in patient care that occurred as a result of differences in outpatient radiograph readings for cases in which the primary care clinician, hypothetically, would not request a second reading by a radiologist.

**Methods:** During 4 months, 1393 pairs of radiographic readings were collected from 9 volunteer primary care practices with 86 clinicians and a second reading by one of 42 radiologists. For 553, hypothetically, the clinician would not request a consultation from a radiologist. Review analysis of the 553 pairs revealed 100 (18.1%) radiographs in which the radiologist’s reading did not agree with the clinician’s reading. Data from the original visit and subsequent related care were abstracted from patient charts and reviewed. Changes in clinical care resulting from the radiologist’s reading were identified.

**Results:** The radiologists’ second reading of these 553 sets of radiographs resulted in 14 (2.5% of 553 and 14% of 100) cases of one or more changes in care. We found 38 documented or presumed changes in care and zero substantial changes in care.

**Conclusions:** Primary care clinicians are able to identify radiographs for which a second reading by a radiologist will not result in substantial changes in care. (J Am Board Fam Pract 2004;17:256–63.)

Thirteen percent of general radiology (plain films) in the United States are performed by family and general practitioners. Analysis of billing data has shown that 42% to 70% of outpatient radiographs are performed and read by nonradiologists. These studies addressed costs and which radiologic procedures were performed by radiologists and nonradiologists. However, none of these studies specifically reported the frequency of outpatient radiographs initially read by the primary care clinician and then referred to a radiologist for a second interpretation.

Halvorsen and Kunian found that 87.3% of Minnesota family physicians had on-site radiographic equipment. Likewise, Smith reported that 76% of Wisconsin family physicians had radiographic equipment in their own offices, and 87% had this equipment in the same building. They also reported that up to 54.2% of family physicians select which radiographs to refer for a radiologist’s reading.

Several studies have compared the concordance of primary care physician interpretation of office radiographs with the interpretations of radiologists and found concordance rates between 87.5% and 93.4%. Studies comparing Emergency Department physicians’ and radiologists’ interpretations of radiographs taken in the Emergency Department have reported concordance rates of 91.8 to 99.3%. A few studies have reported changes in clinical care as a result of discordant readings, but these studies are limited by small numbers of cases, and changes in clinical care were not the primary study aim.

Many radiographs taken in ambulatory settings are either interpreted by the treating clinician alone or with a second reading by a radiologist. We found little evidence in the literature that the second reading adds information necessary for patient care. The primary aim of this study was not to...
determine who made the “correct” reading but to
determine whether adverse patient outcomes might
occur if the primary care clinician did not obtain a
radiology consultation.

Methods
A prospective cohort study was conducted from
April to July 1997. We gathered consecutive pairs
of radiograph readings at 4 urban and 5 rural Wis-
consin Research Network practices that volun-
teered to participate and routinely referred all ra-
diographs for radiology reading. Three sites were
family medicine residency training sites and 6 were
community practices.

During the study period, all child and adult
patients for whom outpatient radiographs were or-
dered by the participating clinicians were eligible.
Informed consent was obtained by the clinic radio-
graph technicians. Patients were not enrolled at
times of high clinic volume when there was insuf-
icient time to obtain consent. The unit of analysis
was the standard “set” of radiographs obtained for
a particular body area. Bilateral radiographs were
counted as 2 “sets” of radiographs.

The participating clinicians were instructed to
interpret their outpatient radiographs in the usual
way. This included their usual process of asking for
the opinion of a colleague or supervisor before the
radiographs were sent for the radiologist’s reading.
Resident physicians are expected to review their
interpretations with faculty preceptors, although
there are occasional instances when this does not
occur. The clinicians were instructed to record
their final interpretation and answer the hypothet-
ical question, “If it were optional, would radiology
over-reading [radiology referral] be requested?” on
the data collection instrument before sending the
radiograph for the second reading. During the
study period, the radiologist’s reports were photo-
copied; the patient’s was name replaced by a nu-
meric identifier and matched up with the study
form with the same identifier.

Each of the 3 family physician authors indepen-
dently reviewed all the pairs of readings for poten-
tial clinically important discordance between the
readings and coded the pairs as concordant or dis-
cordant. The authors coded the pair as discordant
when there was any uncertainty.

Chart reviews were performed by research assis-
tants for all cases when at least 2 of the 3 physician
reviewers coded the interpretation pairs as discor-
dant and the clinician hypothetically would have
declined radiology referral. All materials found in
the chart related to the body area studied with the
radiograph(s) were photocopied starting with the
index visit and ending 6 to 12 months later. These
included office progress notes, emergency depart-
ment records, hospital records, consultations, doc-
umentation of telephone conversations, and radi-
ography or other testing related to the body area
imaged at the index visit. Chart reviews were per-
formed at least 12 months after the index visit.

The chart abstracts were then reviewed independ-
ently by the same 3 family physicians. Any change in
the patient’s care was recorded. We included addi-
tional telephone calls as a change in care in that they
required additional staff and patient time and poten-
tially caused additional anxiety about abnormal re-
results. We assumed that 2 telephone calls occurred
when additional tests occurred as a result of discord-
dance, one to set up the test and one to report the
results. We assumed that 3 telephone calls occurred
when old radiographs were obtained for comparison,
one to determine where the old radiographs were
located, one to obtain the old radiographs, and one to
report the results of the comparison. A substantial
change in care was defined as one likely to cause
harm, such as death, permanent disability, or pro-
longed recovery, if the change had not occurred.
Disagreement between reviewers was resolved by
consensus. Approximately 1 year after the original
chart review, a different research assistant randomly
repeated 9% of the chart reviews and the authors
repeated their evaluation to check for consistency
with the process.

Descriptive analysis summarized radiograph type,
frequency, and hypothetical choice to refer by type.
Inter-rater reliability of discordance between pairs of
readings was measured using the κ statistic.19 For
each body area, the χ² test was used to test for signifi-
cant difference in proportion of agreement between
radiographs when referral would have been hypothet-
ically requested or declined.

Results
A total of 1530 patients had radiographs and inter-
pretations (Figure 1) Seventeen community family
physicians, 3 community surgeons, 16 University of
Wisconsin faculty family physicians, 36 University
of Wisconsin family practice residents, 5 nurse
practitioners, and 9 physician assistants participated in the study. A few initial readings by nonprimary care clinicians were inadvertently included in the data submitted and because we did not collect specialty, degree, or level of training data, it was impossible to remove those readings or analyze those factors. Forty-two radiologists from 3 institutions did at least one reading.

Less than 1% of patients refused participation. The frequency of enrollment varied by site and tended to drop off as the study progressed. Based on the assumption that enrollment was near 100% initially and on the level of drop-off over time, we estimate an average enrollment of 70% of eligible patients. One thousand, three hundred and ninety-three (91%) cases had both readings and an answer to the hypothetical question. Of those 1393 cases, 553 (39.7%) were cases that the clinician hypothetically would have declined referral to a radiologist for a second reading. The agreement between reviewers coding concordance or discordance for the 1393 cases ranged from $\kappa = 0.55$ to 0.60. The overall agreement between any 2 reviewers was 69.4%. For those radiographs that were hypothetically not referred for radiologist reading and classified as discordant by a least 2 reviewers, 100 of 553 (18.1%) had potential clinically important discordance and chart review was performed. After adjusting for differences in types of radiographs being performed and clinician desire for radiology consultation, logistic regression analysis showed one clinic with significantly more discordance between primary care clinician and radiologist (49% versus 28% for the other clinics; odds ratio, 2.07; $P < .01$).

Chest radiographs were the most common radiographs, comprising 29.4% of the total (Table 1). Lower and upper extremity were the next most frequent category, comprising 27.7% and 23.9%, respectively. The frequency at which primary care clinicians would have hypothetically declined radiology referral varied from 50.8% for upper extremity radiograph to 27.1% for chest radiographs.

The concordance between the readings of the primary care clinician and the radiologist for all radiographs was 1010 of 1393 (72.5%) (Table 2). The concordance of radiograph reading for each body area ranged from 52.9% for abdominal radiographs to 72.9% for upper and lower
Concordance for all radiograph readings when radiology referral would have been requested was 66.3%. A higher frequency of concordance was noted when radiology referral would have been declined for all body areas. Concordance varied from 62.5% for spine radiographs to 91.7% of face and head radiographs. Concordance rate for all radiograph readings when radiology referral would have been declined was 81.9%.

There were 100 of 553 (18.1%) cases in which radiology referral would have been declined and in which the readings were discordant (Figure 1). A change in clinical care occurred in 14 of 100 (14%) cases. All cases involved primary care clinicians: 2 physician assistant cases, 3 community family physician cases, 4 resident/faculty family physician cases, and 5 faculty family physician cases. There were 5 cases of definite or possible acute fracture, one case of stress fracture, 5 cases of definite or possible lung nodule, one case of possible mild acromioclavicular subluxation, one case of lumbar spondylolisthesis, and one case of possible pneumonia. There were 4 episodes of discordance related to presence or possible presence of fractures with no follow up visits documented in the chart and unknown outcome (Table 3).

We found 38 documented changes or presumed changes in care (Table 3). The changes included: 23 documented or presumed telephone calls; 5 sets of additional follow-up radiographs; 2 additional office visits; 2 instances in which copies of old comparison radiographs were obtained; 2 repeat radiographs; one cast application; one new prescription; one excuse for gym class; one “possible CT scan” was never scheduled. More than one change in care occurred for some cases. We found zero substantial changes in care or episodes of averted patient harm.

Repeat medical record review of 9 randomly selected cases of the 100 showed complete agreement with the initial review for 8 cases. We discovered documentation of telephone follow-up for one case in which no follow up was found on initial review. This additional case with a change of care was included in the total number reported above.

### Discussion

Use of radiography in the evaluation of ambulatory patients is a common and accepted practice in the United States. Many radiographs taken in primary care clinics are of the lower extremity. Concordance for all radiograph readings when radiology referral would have been requested was 66.3%. A higher frequency of concordance was noted when radiology referral would have been declined for all body areas. Concordance varied from 62.5% for spine radiographs to 91.7% of face and head radiographs. Concordance rate for all radiograph readings when radiology referral would have been declined was 81.9%.

### Table 1. Description of Radiographs Obtained by Primary Care Clinicians and Those for Which Radiologist Referral Reading Was Hypothetically Declined

<table>
<thead>
<tr>
<th>Body Area</th>
<th>% of Total Radiographs</th>
<th>% Referral Declined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest</td>
<td>29.4 (410)</td>
<td>27.1 (111)</td>
</tr>
<tr>
<td>Lower extremity</td>
<td>27.7 (386)</td>
<td>49.0 (189)</td>
</tr>
<tr>
<td>Upper extremity</td>
<td>23.9 (333)</td>
<td>50.8 (169)</td>
</tr>
<tr>
<td>Spine</td>
<td>7.6 (106)</td>
<td>30.2 (32)</td>
</tr>
<tr>
<td>Abdomen</td>
<td>3.9 (55)</td>
<td>38.2 (21)</td>
</tr>
<tr>
<td>Face/head</td>
<td>3.1 (43)</td>
<td>27.9 (12)</td>
</tr>
<tr>
<td>Pelvis</td>
<td>2.7 (37)</td>
<td>29.7 (11)</td>
</tr>
<tr>
<td>Rib/sternum/clavicle</td>
<td>1.7 (23)</td>
<td>34.8 (8)</td>
</tr>
<tr>
<td>Total</td>
<td>100 (1393)</td>
<td>40 (553)</td>
</tr>
</tbody>
</table>

### Table 2. Percentage Concordance between Primary Care Clinician and Radiologist Readings, Referral Hypothetically Requested or Declined

<table>
<thead>
<tr>
<th>Body Area</th>
<th>N</th>
<th>% Concordance All Readings</th>
<th>% Concordance When Referral Requested</th>
<th>% Concordance When Referral Declined</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest</td>
<td>410</td>
<td>67.3 (276/410)</td>
<td>62.5 (187/299)</td>
<td>80.2 (89/111)</td>
<td>.001</td>
</tr>
<tr>
<td>Lower extremity</td>
<td>386</td>
<td>78.0 (301/386)</td>
<td>73.1 (144/197)</td>
<td>83.1 (157/189)</td>
<td>.017</td>
</tr>
<tr>
<td>Upper extremity</td>
<td>333</td>
<td>80.2 (267/333)</td>
<td>72.6 (119/164)</td>
<td>87.6 (148/169)</td>
<td>.001</td>
</tr>
<tr>
<td>Both extremities</td>
<td>719</td>
<td>79.0 (568/719)</td>
<td>72.9 (263/361)</td>
<td>85.2 (303/358)</td>
<td>.001</td>
</tr>
<tr>
<td>Spine</td>
<td>106</td>
<td>61.3 (65/106)</td>
<td>60.8 (45/74)</td>
<td>62.5 (20/32)</td>
<td>.869</td>
</tr>
<tr>
<td>Abdomen</td>
<td>55</td>
<td>58.2 (32/55)</td>
<td>52.9 (18/34)</td>
<td>66.7 (14/21)</td>
<td>.305</td>
</tr>
<tr>
<td>Face/head</td>
<td>43</td>
<td>69.8 (30/43)</td>
<td>61.3 (19/31)</td>
<td>91.7 (11/12)</td>
<td>.010</td>
</tr>
<tr>
<td>Pelvis</td>
<td>37</td>
<td>62.2 (23/37)</td>
<td>57.7 (15/26)</td>
<td>72.7 (8/11)</td>
<td>.364</td>
</tr>
<tr>
<td>Rib/sternum/clavicle</td>
<td>23</td>
<td>69.6 (16/23)</td>
<td>66.7 (10/15)</td>
<td>75.0 (6/8)</td>
<td>.670</td>
</tr>
<tr>
<td>Total</td>
<td>1393</td>
<td>72.5 (1010/1393)</td>
<td>66.3 (557/840)</td>
<td>81.9 (453/553)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

* x² test for difference in proportion of concordance in referral hypothetically requested or declined.
care clinicians’ offices are read a second time by a radiologist, but the exact frequency of referral is unknown.

Our study’s overall concordance rate of 72.5% for radiograph readings was somewhat lower than the 87.5% to 92.4% rates in other published studies. The lower concordance would be expected in that our study counted minor discrepancies as discordant to avoid missing any changes in clinical care, although it is difficult to determine whether previous studies counted minor discrepancies or not. Even with our increased sensitivity for

<table>
<thead>
<tr>
<th>Table 3. Changes in Clinical Care and Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
</tbody>
</table>

PA, posteroanterior.
* Presumed or documented 1 telephone contact for each case.
† Presumed or documented 2 telephone contacts for each case.
‡ Presumed or documented 3 telephone contacts for each case.
§ No follow up documented.

260 JABFP July–August 2004 Vol. 17 No. 4
calling a pair of readings discordant, the range of discrepancies for this study is not substantially different compared with the reported 9.9% to 59% range of discordance among radiologists for radiograph readings.20–27 Bergus et al10 carefully evaluated discordant readings between family physicians and radiologists and found that 35.2% were “interpreted correctly by the family physician.” Our primary aim for this study was not to determine who had the correct reading but rather to determine the effect of the second reading by the radiologist on the care of the patient. If we had attempted to determine who was correct, our rate of accuracy might have been higher and closer to other published rates.

The proportions of hypothetical referrals indicate that the clinicians in our study may have more confidence in their ability to interpret extremity radiographs than other radiographs, similar to results reported by Halvorsen et al.9 Increased confidence in interpreting extremity radiographs is also suggested from our study’s higher concordance with the radiologist’s readings compared with all types of radiographs, also similar to previous studies.8,10

If we assume that changes in care did not occur with the other 453 cases when radiology referral would hypothetically have been declined but the readings agreed, these 14 cases represent 2.5% (14 of 553) of the total. We found zero substantial changes in care or episodes of averted patient harm. These findings are similar to those of other small studies that attempted to evaluate the effect on patient care of discordance between the primary care and radiologist readers.7,8,14,18 The frequency of changes in care that occurred when referral would have been requested remains a substantial but unanswered question. Time and funding restrictions precluded us from exploring this question, but this would be important to address in a future study.

This study has several limitations. This was not a consecutive sample of 100% of the radiograph cases at the participating clinics. We were unable to track the total number of refusals or missed opportunities to enroll patients. However, because there was no systematic method for excluding patients, there is no reason to suspect a selection bias. We were unable to determine the cause of one of 3 family practice teaching sites in which there was an increase in frequency of discordant readings compared with the larger group of clinics. This raised the total frequency of discordances but it did not affect the outcome because none of the 14 cases of change in care occurred at that site.

We did not track the frequency of readings according to training status: (ie, resident, attending physician, nurse practitioner, or physician assistant). Halvorson8 reported that faculty family physicians had only a 3.4% increase (85.2% versus 88.6%) in concordance with radiologists’ interpretations compared with resident family physicians. He also reported an additional 3.5% increase in concordance when faculty and residents did collaborative interpretation. With rare exceptions, the residents in this study did collaborative interpretation, so we do not believe resident participation in this study significantly altered the results. Physician assistants interpreted radiographs in only 2 of the 14 cases of changes in care, so a significant influence on the generalizability of the final results seems unlikely.

Our study protocol did not include permission to contact patients. This would have allowed confirmation of the outcome derived from the chart review, discovered changes in care that the patient received that were not recorded in their primary clinician’s office records, and avoided some or all the 4 “unknown outcome” possible fractures in this study. A larger study should include consent to contact the patients.

The most important limitation is that because the number of discordant readings with changes in care was small, rare events might be missed. Given that zero substantial changes in care or episodes of patient harm occurred, and using Hanley’s method of estimating risk,28 our study showed a 0% to 0.5% chance (95% confidence interval) for these events.

Several factors affect the diagnostic interpretation and recommendations that result from reading radiographs, whether by a primary care clinician or by a radiologist. Although having the clinical history can improve detection of radiographic abnormalities,24,29–32 some studies have shown no benefit.26,33 Human variability also has an effect. For example, radiologists reading identical mammograms twice, 5 months between readings, had the same interpretation for only 84% of the cases.14

Another issue is the context bias that is unavoidable for both the primary care clinician and the radiologist. Context bias, as described by Egglin...
and Feinstein,\textsuperscript{35} refers to the effect of higher prevalence of disease on interpretation of diagnostic radiographs. Radiographs from a population with a low frequency of disease intermingled with radiographs with a high frequency of disease are more likely to be interpreted as abnormal. Primary care clinicians see a population in their offices with lower risks of most radiologic abnormalities than the hospital and emergency department population that generates the majority of the radiologists' films to interpret. This difference in disease frequency will influence the radiologist to interpret equivocal findings as abnormal. This context bias, combined with the history and physical examination before radiographs are obtained and again after, if necessary, allows the primary care clinician to dismiss findings that are incidental.

With an estimated $38.5 billion spent in the United States on all radiologic studies in 1997\textsuperscript{16} and growing every year, a larger multiregional, multicenter primary care study to assess the value or detriment of the second reading of outpatient radiographs seems warranted. A future study would include methods beyond chart review to determine the final clinical outcome and would be sufficiently powered to discover relatively rare events. In addition to more research on context bias in the ambulatory setting, a larger study could result in guidelines for requesting radiology referral for ambulatory radiographs.

This is the first study to specifically address the effect of a second reading of office radiographs on the care of patients. We found little added benefit with a very small frequency of any change in clinical care. The majority of the changes were episodes of unnecessary additional radiologic procedures, administrative effort, or follow up care, without substantive improvement in the clinical care of the patient. We conclude that primary care clinicians are able to identify those radiographs for which a second reading by a radiologist will not result in substantial changes in care or episodes of patient harm.

Increased interpretation of office radiographs by family physicians has financial implications but not cost savings. Any physician may charge for the technical component (taking the picture) and interpretation as long as a separate written report is generated that addresses the findings, relevant clinical issues, and comparative data (when available), similar to the method of billing for electrocardiograms. Family physicians charging for more of their interpretations of radiographs will shift reimbursement from radiologists to family physicians, but the cost to the health care system remains the same. Liability concerns may drive some health care organizations to demand that all ambulatory radiographs be read by a radiologist, but the literature does not support such a policy.

There is an increasing shortage of radiologists in the United States.\textsuperscript{37–39} Primary care clinicians selecting which radiographs to send for second reading would free up radiologists' time for interpretation of more complex radiographs and radiologic interventions. Moreover, up to 45% of rural clinicians do not have daily access to a radiologist,\textsuperscript{5} and up to 73% of outpatient chest, spine, pelvic, and extremity radiographs are not read a second time by a radiologist.\textsuperscript{4} This study's results will give them some measure of confidence that important missed diagnoses are, at the worst, very rare events. Universal second readings are not warranted based on our findings.

We acknowledge the important contributions of the Wisconsin Research Network (WReN) clinicians, their radiography technicians and the office staff, who gathered the data. We thank Linda Manwell, RN, Mary Stone for secretarial support, and Pamela Wiesen, MBA, for project support.

References
8. Halvorsen JG, Kunian A, Gjerdingen D, Connolly J,


